James P. Scanlan
2638 39th Street, N.W.
Washington, D.C. 20007
(202) 663-4791 (O)
(202) 337-3927 (H)

About 5750 Words

Direct Regression and Reverse Regresssion: Contrasting Ways of Looking at the Wrong Information

James P. Scanlan
Equal Employment Opportunity Commission,
Washington, D.C. 20507

ABSTRACT: Labor economists and statisticians have debated whether direct regression or reverse regression is more useful for determing whether employment discrimination has occurred. The former technique tends to find discrimination against the less-qualified group while the latter technique tends to find discrimination against the more-qualified group. Neither side to the controversy, however, has recognized that, in most of the contexts where regression techniques are used to prove employment discrimination, both direct and reverse regression approaches are fundamentally flawed because they fail to consider the entire universe of persons subject to the decision-making process at issue.

KEY WORDS: Regression techniques; Employment discrimination; Placement discrimination; Job segregation; Pay discrimination; Initial assignment.

1. INTRODUCTION

In the early 1980s, this journal featured a controversy concerning whether direct regression or reverse regression is the more suitable technique for examining the fairness of employment practices. Direct regression has been criticized for its bias in the direction of indicating discrimination against the less-qualified group, while reverse regression has been criticized for its bias in the direction of indicating discrimination against the more-qualified group. Although the issue has never been definitively resolved, direct regression remains the procedure ordinarily utilized by both sides in employment discrimination litigation. In most circumstances in which they are used, however, both procedures are fundamentally flawed techniques for identifying discrimination. The flaw lies not simply in the biases that have commonly been noted in each technique, but in the failure of either approach to examine the full universe of persons subject to the decision-making process at issue.

Conway and Roberts first brought the debate over direct and reverse regression into prominence here in 1983, relying on hypothetical data from a large bank. The data apparently was drawn from the Harris Bank case, a proceeding before a Department of Labor Administrative Law Judge, in which one of the authors had testified. (See Blattenberger and Michelson 1984, Conway and Roberts 1986). The case had involved allegations of discrimination against minorities and women with respect to salary and placement.

To illustrate the differing concepts of fairness that were the focus of direct and reverse regression, Conway and Roberts presented a hypothetical distribution of the placement of men and women with varying educational attainments in professional and clerical jobs. That hypothetical is shown in Table 1.

Based on the data in Table 1, Conway and Roberts described two concepts of fairness. What Conway and Roberts termed "Fairness 1"--the focus of direct regression--concerned the issue of whether men and women of equal measured qualifications were placed in the same jobs; the presence or absence of Fairness 1 could be determined by comparisons of horizontal percentages. "Fairness 2"--the focus of reverse regression--concerned the issue of whether men and women in the same jobs were equally qualified; the presence or absence of Fairness 2 could be determined by examination of vertical percentages.

Table 2 provides the horizontal percentages that reflect the direct regression approach to identifying discrimination. Examining percentages in Table 2, we would conclude that Fairness 1 is violated (a situation Conway and Roberts also described as "Type 1 Unfairness"). The percentages show that on average women are not placed in professional jobs at the same rates as equally qualified men: 42.5 percent of college-educated men are placed in professional positions compared with only 20 percent of college-educated women. From the Fairness 1 perspective, then, the placement process in unfair to women.

Table 3 provides the vertical percentages that reflect the reverse regression approach to identifying discrimination. Examining the percentages in Table 3, we would conclude that Fairness 2 is violated ("Type 2 Unfairness"). From this perspective, however, the unfairness is to men. On average women who are placed in professional jobs are less qualified than men placed in such jobs. While 52.8 percent of men in professional jobs have MBAs, none of the women do. Women in clerical positions are also less qualified than men. While 83.3 percent of clerical women have only a high school diploma, 62.5 percent of clerical men have only a high school diploma.

Conway and Roberts explained that these paradoxical results stem from the fact that, on average, men in the company have higher qualifications than women. They also pointed out that because of that circumstance, even extreme Type 1 Unfairness to women could coexist with Type 2 Unfairness to men. And they noted that the hypothetical is realistic in many applications, with direct regression studies that have been performed tending in fact to find salary shortfalls for women at the same time that they find female shortfalls in job qualifications.

A number of economists and statisticians responded here to Conway and Roberts' 1983 article, criticizing the reverse regression approach for several reasons (Ferber and Green 1984, Goldberger 1984, Greene 1984, Michelson and Blattenberger 1984, and Miller 1984). Conway and Roberts (1984) replied to those criticisms and pointed out as well that both reverse and direct regression had information to offer in evaluating the fairness of employment practices. Others participated in the debate in other places, usually taking issue with the reverse regression approach (Weisberg and Tomberlin 1983 and Ash 1986).

I attempt to show below why these treatments have generally failed to focus on the fundamental error of reverse regression, and have failed to recognize that the same error renders direct regression an equally flawed method of proving discrimination in most of the contexts in which it has been employed.

2. THE FALLACY OF REVERSE REGRESSION

The reason that reverse regression is a flawed technique can be simply stated: A group that is on average less qualified among applicants for a position will comprise a larger proportion of the less-qualified applicants for the position than of the more-qualified applicants for the position; accordingly, the less-qualified group will tend to be less-qualified among the persons selected for the position when the selection process is completely unbiased and often even when there exists substantial discrimination against the less-qualified group. Regardless of this bias, however, the technique is fundamentally flawed because it attempts to draw conclusions about the selection process without considering the entire universe of persons subject to that process.

2.1 Conway and Robert's Treatment of Finkelstein's Hypotheticals

To a large extent the fallacy of reverse regression was illustrated by Conway and Roberts themselves in treating objections to reverse regression that had been raised by Finkelstein (1980) and Weisberg and Tomberlin (1983). Conway and Roberts first presented tables illustrating two hypotheticals proposed by Finkelstein. In these hypotheticals Finkelstein had posited a promotional situation where women were less qualified among the persons seeking to be promoted from clerical to professional positions. In his first hypothetical, set out here in Table 4, Finkelstein posited a case where, among both persons with college degrees and without college degrees, women were promoted at lower rates than men, with the result that after promotion men and women in professional positions were equally likely to have college degrees. As Conway and Roberts characterized it, Type 1 Unfairness (toward women) had led to Type 2 Fairness.

In Finkelstein's second hypothetical, set out here in Table 5, he posited the case where, among both persons with college degrees and without college degrees, men and women were equally likely to be promoted, with the result that among persons promoted to professional positions men were more likely than women to have college degrees. In Conway and Robert's terms, Type 1 Fairness had led to Type 2 Unfairness (toward men).

Conway and Roberts maintained that Finkelstein's hypotheticals did not disprove the validity of reverse regression, in their view, because the initial situation was unfair to men in the Type 2 sense. As Conway and Roberts put it, the Type 1 Unfairness (to women) in promotions simply corrected the initial Type 2 Unfairness (to men), reasoning that Conway and Roberts argued was akin to the civil rights position that differential treatment may be necessary to correct past discrimination.

Yet, analyses of employment practices for purposes of determining whether discrimination exists must focus on whether similarly-situated individuals are treated equally. Affirmative action considerations and other arguments about correcting existing inequities enter the picture only as possible justifications for treating people differently. In Finkelstein's first hypothetical, men and women were not treated equally with respect to promotions. Reverse regression, however, indicated that there was no discrimination. Indeed, as shown in Table 6, with a slight modification to Finkelstein's first hypothetical, the situation is created where women are not as likely to be promoted as similarly-situated men, but reverse regression still indicates Type 2 Unfairness against men.

Finkelstein's second hypothetical could be adjusted to the same effect. Actually, Finkelstein made his point unduly complicated by presenting different pre-promotion data in his two hypotheticals. One set of pre-promotion data would serve as a basis for revealing all three phenomena that expose the fallacy of reverse regression: (1) Type 1 Unfairness against the less-qualified group can yield Type 2 Fairness; (2) Type 1 Fairness will necessarily yield Type 2 Unfairness against the more-qualified group; and (3) Type 1 Unfairness against the less-qualified group can still yield Type 2 Unfairness against the more-qualified group.

In explaining away Finkelstein's hypotheticals, Conway and Roberts confused justifications for the differential treatment of similarly qualified men and women with the issue of whether, and at what points in the employment process, discrimination occurred.

Finkelstein's hypotheticals somewhat lent themselves to Conway and Roberts' misfocusing of the issue because the hypotheticals involved a pre-promotion situation that Conway and Roberts could characterize as infected by Type 2 Unfairness. As I will show below, that misfocus could have been avoided by reference to data on applicants and hires.

2.2 Conway and Roberts' Treatment of Weisberg and Tomberlin

Conway and Roberts also addressed another effective refutation of reverse regression, which had been presented by Weisberg and Toberlin (1983). Weisberg and Tomberlin presented a hypothetical distribution of male and female test scores where males had average higher scores. They then showed that selection of the six highest scorers from the total pool would yield a group of hires where the average male score was higher than the average female score. Conway and Roberts responded by arguing that, assuming that test scores are accurate predictors of performance, Fairness 2 is violated because the men are not being compensated according to their productivity.

Conway and Roberts' response, however, merely raises a question about whether or not it is fair to pay the same salary to persons who are in the same job even though some are more productive than others. People might differ on the answer to that question, though few would maintain that when an employer pays one set salary for a particular job it engages in anything called discrimination.

Conway and Roberts' response, however, wholly fails to address Weisberg and Tomberlin's objection that, when the issue is whether the selection process treated people equally, reverse regression would give the misleading impression that the employer had treated men unfairly by imposing more exacting standards on them when in fact the selection process was completely nondiscriminatory. And, as previously explained with respect to Finkelstein's hypothetical, in many circumstances reverse regression could also create the impression that the employer had favored the less-qualified group when in fact it had favored the more-qualified group.

2.3 Reverse Regression in the *Sears* Case

When the inquiry is properly focused upon the issue of whether the selection process discriminates on the basis of race or gender, the arguments presented by both Finkelstein and Weisberg and Tomberlin should be read to conclusively illustrate the error of reverse regression. A clearer illustration of that error, however, may be found in actual data on applicants and hires from the widely-publicized case of EEOC v. Sears, Roebuck and Co (1986, 1988). The EEOC (for which I was counsel in the case) had contended that Sears engaged in hiring discrimination against women with respect to commission sales positions. Although the term "reverse regression" was not used in the case, Sears presented a variety of data showing that with respect to numerous characteristics, including actual post-hire sales performance, men hired into commission sales jobs were on average more qualified than women hired into such jobs. Sears contended that this was evidence that it had actually favored women in hiring for commission sales positions.

The EEOC, however, presented numerous exhibits showing both hypothetically, and with reference to the data in the case, that because women were less-qualified among the applicants seeking jobs at Sears, they would tend to be less-qualified among the hires even when they were less likely to be selected than similarly-qualified men. Some of that evidence is present in Table 7, which shows that women were less likely to be selected than men both from among applicants with and without prior experience. Women were 40 percent of applicants with experience, but only 20 percent of hires with experience; they were 70 percent of applicants without experience, but only 30 percent of hires without experience. Nevertheless, women who were hired were less likely to have experience than men who were hired. Only 40 percent (10 of 25) of women who were hired had experience compared with 53 percent (40 of 75) of men.

In should be kept in mind that evidence like that in Table 7, while suggestive that Sears discriminated against women, was not definitive evidence of such discrimination. Other factors, including that, on average, experienced men had more experience and better experience than experienced women, could provide nondiscriminatory explanations for the disparity. The table does show, however, why the fact that men who were hired were more likely to have experience than women who were hired does not indicate any form of unfairness toward men. It also does not negate the possibility of unfairness against women with respect to the only type of fairness that can have meaning in the evaluation of a hiring process--i.e., whether men and women with similar qualifications have the same chance of being hired.

As it happens, although it was a minor aspect of the decisions in Sears' favor, both the district court and the court of appeals pointed to the greater qualifications of male hires as evidence that Sears had been favoring women, rather than discriminating against them. Only Judge Richard D. Cudahy, dissenting from the court of appeals decision, recognized that the group that was more qualified among the applicants would be more qualified among the hires regardless of whether there was discrimination against the less-qualified group.

Nevertheless, I submit, evidence like that presented in Table 7 demonstrates, in unassailable mathematical terms, that in Sears and all other cases, given evidence that one group is more qualified among applicants, the fact that that group is more qualified among hires is not inconsistent with discrimination against the less-qualified group.

And as a rule, of course, the group that is alleged to have been discriminated against in a selection process will tend to be the group that is less qualified among applicants for the positions in question. This frequently is the reason that employers will discriminate against them (Phelps 1972, Posner 1987, Scanlan 1988). But even if there merely exists the possibility that one group is less qualified among applicants than another, it is not reasonable to infer the presence or absence of discrimination through a reverse regression technique that looks solely at the qualifications of the persons who are actually hired.

Fortunately, the reverse regression technique has not become a common approach to the defense of employment discrimination litigation. But its essential elements--including its essential flaw--do appear in popular, and even occasionally in scholarly, discussion of a variety of issues. For example, the higher average qualifications of Asians admitted to various universities is sometimes cited as strong evidence of discrimination against Asian applicants (Washington Post 1988). Similarly, the superior performance of black professional athletes is frequently cited as evidence that blacks are discriminated against in the selection process (Lapchick and Slaughter 1989, p. 58, Kahn 1991). It generally goes unrecognized that when a group that is more qualified among persons seeking a position, the group usually will be more qualified among the persons selected whether or not there is discrimination against it. Though it is entirely conceivable that the suspected discrimination exists in either or both of these contexts, data solely on the qualifications of persons who have been selected is not probative of such discrimination (Scanlan 1988, 1991).

3. THE FALLACY OF DIRECT REGRESSION

The fundamental error of direct regression in most of its applications in employment discrimination contexts is the same as the fundamental error of reverse regression. That is, the procedure treats persons actually hired as if they comprised the entire population subject to the decision-making process as issue.

This is not the same objection to direct regression raised by Roberts (1979) and Dempster (1984) in arguing the superiority of reverse regression, or that was acknowledged, for example, by Ash (1986) while criticizing reverse regression. Such objections rest essentially on the fact that imperfections in measuring attributes tend systematically to bias results in the direction of finding discrimination against the less-qualified group; but they do not involve challenges to the appropriateness of conducting the analysis solely on persons hired. The tendency of direct regression to find discrimination against the less-qualified group whenever unmeasured attributes are positively correlated with measured attributes, to be sure, is a common weakness of direct regression techniques. Notwithstanding those weaknesses, however, direct regression performed on the correct population may still be probative. In such circumstances, the question simply will be whether, all things considered, it is more likely than not that the weaknesses make enough of a difference to alter the conclusion.

The failure to consider the entire population subject to the decision-making process is a more fundamental error.

3.1 Illustration of the Fundamental Error

The fundamental error of failing to consider the entire universe of person subject to the selection process is illustrated by the hypothetical set out in Table 8, which is drawn in part from the data that Conway and Roberts used to illustrate how direct regression indicated discriminatory placement of college-educated women in the Harris Bank case. As already shown in Table 2, college-educated women were disproportionately placed in clerical positions (80.0%) relative to college-educated men (57.5%). To argue that this somehow suggests discrimination, however, is to ignore the hiring process that brought these individuals into the employer's work force and the possibility (or probability) that the gender composition of the applicant pools for each job differed.

Suppose that in the pool of college-educated applicants seeking jobs with the employer, college-educated women comprised a higher percentage of the persons seeking clerical positions than of persons seeking seeking professional jobs. Table 8 posits that women made up 20% of college-educated applicants interested in professional positions, while they comprised 40% of applicants interested in clerical positions. As shown in the table, if such was the gender makeup of the applicant pool, then essentially equal hiring of male and female college-educated applicants would yield the placement patterns that direct regression would interpret as evidence of placement discrimination.

It should be borne in mind that the differing gender compositions of applicants for the two jobs need not suggest that any college-educated women are more interested in clerical jobs than in professional jobs. Rather, it merely suggests that women will accept clerical jobs, if offered, at a higher rate of men. The applicants pool for the clerical pool, indeed, could be comprised entirely of persons also seeking (and preferring) professional positions. But, if college-eduated women seeking professional positions were about 2.7 times as likely as college-educated men seeking those positions to be also willing to take clerical positions if offered, the applicant pool would be that reflected in Table 8.

There is no point in wondering whether this particular gender difference in willingness to take the clerical job seems high or low. It happens to be the figure that would explain the seemingly disparate placement patterns of college-educated men and women in the Harris Bank case as flowing from nondiscriminatory hiring practices (assuming, though unnecessarily, that all college-educated persons willing to take clerical positions were also seeking professional positions). Other hypotheticals could be used employing varying assumptions about the overlap of the pools of persons seeking/willing to accept better and poorer jobs. The essential point, however, is simply that, as long as there is some gender difference in willingness of persons of given qualifications to accept certain types of jobs, direct regression will indicate placement discrimination even when the employer has treated applicants from the two groups equally.

It should also be borne in mind that it does not matter that ~~the~~ a regression analysis in some manner attempts to adjust for the "job-applied-for" indicated on the applications of persons actually hired. Critical for determining whether similarly-qualified men and women seeking the better job had the same chance of being hired is a comparison between the gender composition of hires into the better job and the gender composition of the applicants seeking the better job. Persons hired into the poorer job who also were seeking the better job are part of the pool for the better job. Usually they are a small part of that pool; but, in any case, it is essential to know the composition of the pool including the usually large number of persons who were not hired at all.

3.2 Anomalies of Internal Comparisons.

Regression techniques where the results are at all affected by the initial hiring decision are but seemingly sophisticated variations on the commonplace naive efforts to find discriminatory exclusion from certain jobs based on the perceived overrepresentation in poorer jobs. I have catalogued those approaches elsewhere (Scanlan 1988, 1989a, 1989b).

Such approaches contain anomalies apart from the tendency to find discrimination wherever one group is more willing to accept the poorer job. Consider the situations depicted in Tables 9 through 12, which involve employers who draw from the same labor markets for the same types of jobs. That is, we assume that the gender composition of applicants for the High Job is the same for each employer and the gender composition of applicants for the Low Job is the same for each employer. Since we do not know what the gender compositions for either job is, we cannot determine whether any of the employers actually discriminates against women. What we consider here are the varying conclusions we might draw about each employer relying solely on the distributions of persons actually hired.

In the hypothetical reflected in Table 9, among persons hired by Employer One, men are 1.7 times as likely as women to be employed in the High Job. Suppose, however, that Employer Two draws from the same labor market for the same types of jobs and hires women for each type of job at the same percentage rate as Employer

One; but Employer Two has twice as many Low Jobs as Employer One. The breakdown of Employer Two's male and female employees is shown in Table 10.

As already noted, since we do no know the gender composition of the applicants for each job, we cannot know whether either employer discriminates against women in hiring into either job. But since we assume that the gender composition of applicants is the same for each employer, we can conclude that the employers have identical practices with respect to hiring women into each job. Yet, examination of the distributions reveals that among persons hired by the Employer Two, men are twice as likely as women to be hired into High Jobs, compared with 1.7 times for Employer One.

Another way of evaluating the data is through the Index of Occupational Segregation--i.e., the proportion of employees of either gender who would have to change jobs in order to equalize the distributions. This indicator is more often used for appraising job segregation beyond the individual employer (Reskin and Hartman 1986, Baker 1988, Fields and Wolff 1991), but it has also been used for analysis of discriminatory assignment patterns within individual firms (Bielby and Barron 1984, Hartman 1987, Cassel, Director, and Doctors

1975, Osterman 1979). Though, as indicated, Employers One and Two actually treat women identically with respect to hiring for both jobs, Employer One has a Segregation Index of .24, while Employer 2 has a segregation index of .20. Thus, by this standard, Employer One is more segregated than Employer Two.

A further anomaly is reflected in Table 11, which presents the employees of hypothetical Employer Three, who hires the same number of persons for the same types of jobs as Employer One, and who has the same gender composition of applicants for its High and Low Jobs that Employer One has. Women comprise a smaller proportion of the persons Employer Three hires for each category of jobs than was the case for Employers One and Two. Although Employer Three thus treats female applicants less favorably than Employer One, the disparity between the rates at which male and female employees are in High Jobs is lower for Employer Three than for Employer One; Employer Three's male employees are 1.6 as likely as to be in High Jobs as its female employees, compared with 1.7 times for Employer One. Employer Three also has a lower Segregation Index (.20) than Employer One.

Finally, Table 12 shows hypothetical Employer Four, who hires the same numbers of employees for each job as

Employers One and Three, and who has the same gender composition of applicants for each job. Employer Four, however, while hiring women at the same rate as Employer Three for the High Job, also hires women at that rate for the Low Job. Though Employer Four treats women even less favorably than Employer Three, there is no disparity in the rates at which its male and female hires are placed in the High Job, and its segregation index is zero, indicating no segregation whatever.

All these comparisons, however, like any analysis that fails to include a critical element, are ultimately meaningless. And no matter how sophisticated a regression technique may appear, it is no less meaningless if it simlarly ignores the gender composition of the person seeking each type of job.

4. OBJECTIONS

Of all those writing about regression, only Weisberg and Tomberlin (1983) appear to recognize that the real universe of interest in any case affected by initial hiring decisions includes the rejected applicants. Weisberg and Tomberlin nevertheless appear to regard direct regression as an acceptable approach, evidently based on the view that the persons actually hired can provide a reasonable proxy for the entire pool of applicants.

Yet there is no reason to believe that persons hired in fact would likely mirror the pool of applicants. We have already seen in Tables 9 through 12 how employers with identical applicant pools would for purposes of a direct regression analysis be deemed to have an expected female representation in the better jobs (assuming nondiscriminatory placement) of 30 perecent, 33 percent, 15 percent, and 10 percent--i.e., the female representation of each employer's total work force.

Some might argue that it is improper to assume that similarly-qualified persons from different groups would differ with respect to willingness to accept the poorer job. Yet, the very fact that is believed that there exists discrimination against minorities and women pervasive enough to systematically diminish their employment opportunities powerfully suggests reason for them to accept less desirable positions than similarly-qualified whites and men.

In the case of women, there also often are constraints upon job choices resulting from the fact that (even in two-parent households) women bear a disproportionate share of child care responsibilities and the fact that they tend to be less mobile, at a minimum, because more often than men they have an employed spouse, usually one earning a greater income than they do. Such constraints would naturally tend to make women readier to accept the less-than-preferred position with a particular employer.

There is another reason to expect minorities and women to accept poorer jobs than whites and men with similar measured characteristics that is related to, but in a significant way different from, Conway and Roberts' basic objection to direct regression. As already explained, that objection rests on the probable correlation of unmeasured productivity-related characteristics with measured productivity-related characteristics. As also noted, this is a reasonable objection even to a regression technique that focuses on the actual universe of persons subject to the decision-making process at issue. But it is also an objection, as pointed out by Michelson and Blattenberger (1984), that assumes, perhaps unjustifiably, that management knows but does not record the unmeasured productivity-related characteristics.

The point raised by Michelson and Blattenberger does not apply in the same way--and is probably less valid--with regard to applicants' own knowledge of characteristics beyond those listed in applications and other material from which the analyst would derive the measured characteristics for a regression analysis. Thus, to the extent that applicants from the less-qualified group tend to have lower actual qualifications than persons from the more-qualified group who have similar measured qualifications, the applicants' knowledge of these differences would tend to lead members of the less-qualified group to believe that they have a weaker bargaining position. That belief would be attended by a greater readiness to accept the poorer job.

In any case, whatever the causes of the different gender composition of persons of any skill level willing to accept different jobs (or certain salaries), and whatever the extent of the differences, the fact is that the gender composition of the pool for each job is a critical element in analyzing whether the employer treats equally-qualified people without regard to demographic characteristics.

Some might further object that to fail to permit demonstrations of discrimination through the analysis solely of persons hired would permit some discrimination to go undetected, a result that would be especially unfair where the employer has itself destroyed the applications of rejected applicants necessary for a hiring analysis. Yet, though it is regrettable that some discrimination may go undetected where there is insufficient data available for a competent analysis, this is hardly an argument for permitting incompetent analyses that do not offer a reliable approach for estimating the missing information. That an employer has destroyed applications in violation of statutes or regulations may provide reason for courts to be less demanding of the precision with which a plaintiff estimates factors critical to its case, or to impose more onerous rebuttal burdens of the employer. But it would not be reason for treating persons actually hired as if they offer a reasonable proxy for persons who applied--an approach incidentally that would lead to findings of discrimination against employers in Tables 9, 10, and 11, but that would exonerate the employer in Table 12. Unless, it can be shown that the characteristics of hires, including their gender composition, reasonably approximate the characteristics of applicants, the former is no a more a substitute for the latter than data created from thin air. It is doubtful that such a showing often can be plausibly made.

5. IMPLICATIONS

The essential error of the direct regression technique described above applies wherever initial selection affects the regression results and the characteristics of persons not selected are not somehow incorporated into the analysis. Most obviously, direct regression is not properly used in cases challenging initial assignment, as in, for example, Greenspan v. Automobile Club of Michigan (1980) and the much-discussed case of Vuyanich v. Republic National Bank (1980). But that error is no less present in the analysis used in the many cases where plaintiffs seek to prove discrimination by comparing the present status or pay of persons of similar qualifications when that status is a result of initial placement as well as promotion.

The error also undermines any pay discrimination analysis that does not separate out initial starting salary. For, just as there are different pools of interested applicants for different jobs, there are different pools for the same job at different starting salaries. The same factors that may incline minorities or women to accept lower-level jobs than whites or men of similar measured qualifications would incline them to accept the same jobs at lower salaries.

The rejection of direct regression may seem most troubling in the context where pay is the sole issue. In a case that is initially perceived as involving placement discrimination, one can at least easily enough conceptualize the correct mode of proof for establishing such discrimination as may have occurred at the point of selection/non-selection. That is, the applicants seeking the better job (including those also willing to accept the poorer job) plainly enough comprise the relevant pool for the better job. When employers keep the information that they are supposed to be keep, the essentials of the analysis should be available.

But when the disparities in current status materially result from initial differences in pay even for the same job, it is difficult to conceive of a feasible analysis that could effectively distinguish whether a particular salary structure resulted from race/gender-neutral bargaining or from the employer's use of bargaining approaches that varied depending on the race or gender of the particular applicant. Yet, the very fact that it is so difficult to frame an appropriate approach for analyzing disparities in initial salaries is but further reason to require that a plaintiff's analysis separate out such part of any overall pay disparity that is attributable to starting salaries. Only when the inquiry is carefully focused on the mechanisms that actually may be infected by discrimination is it likely that techniques will be developed for revealing that discrimination.

In any case, however, the fact that discrimination is impossible to prove without resort to methods that are not actually probative of discrimination is not sufficient reason for the acceptance of such methods.

--End of Text--

This article was written in the author's private capacity. No official support of the United States Equal Employment Opportunity Commission or any other federal agency is intended or should be inferred.

REFERENCES

Ash, A. S. (1986), "The Perverse Logic of Reverse Regression," in Statistical Methods in Discrimination Litigation," eds. D.H. Kaye and M. Aikin, New York: Marcel Decker, Inc., pp. 85-105.

Baker, I. (1988), "Women's Employment in Comparative Perspective," in Feminization of the Labor Force: Paradoxes and Promises, eds. J. Jensen, E. Hagen, and C. Reddy, New York: Oxford University Press, pp. 17-43.

Bielby, W. T., and Barron, J. N. (1984), "A Woman's Place is With Other Women: Sex Segregation Within Organizations," in Sex Segregation in the Workplace: Trends, Explanations, Remedies, ed. B. F. Reskin, Washington, D..C.: National Academy Press, pp. 27-55.

Cassel, F. H., Director, S. M., and Doctors, S. I. (1975), "Discrimination Within Internal Labor Markets," Industrial Relations, 14, 337-44.

Conway, D. A., and Roberts, H. V. (1983), "Reverse Regression, Fairness, and Employment Discrimination," Journal of Business and Economic Statistics, 1, 75-85.

Conway, D. A., and Roberts, H. V. (1984), Rejoinder to Comments on "Reverse Regression, Fairness, and Employment Discrimination," Journal of Business and Economic Statistics, 1, 75-85.

Conway, D. A., and Roberts, H. V. (1986), "Regression Analysis in Employment Discrimination Cases," in Statistics and the Law, eds. M. H. DeGroot, S . E. Fienberg, and J. B. Kadane, New York: John Wiley, pp. 107-68.

Dempster, A. P. (1984), "Alternative Methods for Inferring Employment Discrimination from Statistical Data," W.G. Cochran's Impact on Statistics, eds. P. S. R. S. Rao and J. Sedransk, New York: John Wiley, pp. 309-330.

Ferber, M. A., and Green, C. A. (1984), "What Kind of Fairness is Fair? A Comment on Conway and Roberts," Journal of Business and Economic Statistics, 2, 111-113.

Fields, J., and Wolff, E. N. (1991), "The Decline of Sex Segregation and the Wage Gap, 1970-80," Journal of Human Resources, 26(4), pp. 608-22.

Finkelstein, M. O. (1980), "The Judicial Reception of Multiple Regression Models in Legal Proceedings," Columbia Law Review, 80, 737-754.

Goldberger, A. S. (1984), "Redirecting Reverse Regression," Journal of Business and Economic Statistics, 2, 114-116.

Greene, W. H. (1984), "Reverse Regression: The Algebra of Discrimination," Journal of Business and Economic Statistics, 2, 117-120.

Hartmann, H. I. (1987), "Internal Labor Markets and Gender: A Case Study in Promotion," in Gender in the Workplace. eds. C. Brown and J. A. Pechman, Washington, D.C.: Brookings, pp. 59-106.

Kahn, L. M. (1991), "Discrimination in Professional Sports: A Survey of the Literature," Industrial and Labor Relations Review, 44, 395-418.

Lapchik, R. E., and Slaughter, J. B. (1989), The Rules of the Game: Ethics in College Sport, New York: Macmillan.

Michelson, S., and Blattenberger, G. (1984), "Reverse Regression and Employment Discrimination," Journal of Business and Economic Statistics, 2, 121-122.

Miller, J. J. (1984), "Some Observations, a Suggestion, and Some Comments on Conway-Roberts Article," Journal of Business and Economic Statistics, 2, 123-125.

Osterman, P. (1979), "Sex Discrimination in Professional Employment: A Case Study," Industrial and Labor Relations Review, 14, 451-64.

Phelps, E. S. (1972), "The Statistical Theory of Racism and Sexism," American Economic Review 62, 659-61.

Posner, R. A. (1987), "The Efficiency and Efficacy of Title VII," University of Pennsylvania Law Review, 136, 513-21.

Reskin, B. F., and Hartman, H. I. (eds.) (1986), Women's Work, Men's Work: Sex Segregation in the Workplace, Washington, D.C.: National Academy Press.

Scanlan, J. P. (1988), "Illusions of Job Segregation," The Public Interest, 93, 54-69.

_____________ (1989a), "Indiscriminate Reading of Statistics Can 'Prove' Bias Where None Exists, Manhattan Lawyer, April 25-May 1, p. 11.

_____________ (1989b), "Are Bias Statistics Nonsense?" Legal Times, May 8, p. 22.

_____________ (1991), "The Perils of Provocative Statistics," The Public Interest, 102, 3-14.

Weisberg, H. I., and Tomberlin, T. J. (1983), "Statistical Evidence in Employment Discrimination Cases," Sociological Methods and Research, 11, 386-406.

The Washington Post (1988), "The Specter of Quotas," December 17, p. A18.

CASES

EEOC v. Sears, Roebuck and Co., 628 F.Supp. 1264 (N.D. Ill. 1986), affirmed, 839 F.2d 302 (7th Cir. 1988).

Greenspan v. Automobile Club of Michigan, 495 F.Supp. 1021 (E.D. Mich. 1980).

Office of Federal Contract Compliance Programs, United States Department of Labor v. Harris Trust and Savings Bank, Case No. 78-OFCCP-2

Vuyanich v. Republic National Bank of Dallas, 505 F.Supp. 224 (N.D. Tex. 1980)

TABLES

Table 1. Conway and Roberts' Hypothetical Distribution of Male and Female Employees by Education and Entering Job

	Male			Female		
Eduction	Clerical	Professional	Total	Clerical	Professional	Total
High School	40	0	40	80	0	80
College	23	17	40	16	4	20
MBA	1	19	20	0	0	0
Total	64	36	100	96	4	100

Table 2. Conditional Distribution of Entering Job Given Education for Male and Female Employees--Conway and Roberts' "Type 1 Unfairness"

	Male			Female		
Education	Clerical	Professional	Total	Clerical	Professional	Total
High School	100.0	.0	100.0	100.0	.0	100.0
College	57.5	42.5	100.0	80.0	20.0	100.0
MBA	5.0	95.0	100.0	.0	.0	.0
Total	64.0	36.0	100.0	96.0	4.0	100.0

Table 3. Conditional Distribution of Education Given Entering Job for Male and Female Employees--Conway and Roberts' "Type 2 Unfairness"

	Male			Female		
Education	Clerical	Professional	Total	Clerical	Professional	Total
High School	62.5	.0	40.0	83.3	.0	80.0
College	35.6	47.2	40.0	16.7	100.0	20.0
MBA	1.6	52.8	20.0	.0	.0	.0
Total	100.0	100.0	100.0	100.0	100.0	100.0

Table 4. Distribution of Male and Female Employees by Education and Job--First Finkelstein Hypothetical (Unfairness 1 Leads to Fairness 2)

Education	Male			Female		
	Clerical	Professional	Total	Clerical	Professional	Total
	a. Before Promotion					
High School	20	0	20	25	0	25
College	25	0	25	20	0	20
Total	45	0	45	45	0	45
	b. After Promotion					
High School	10	10	20	20	5	25
College	5	20	25	10	10	20
Total	15	30	45	30	15	45

Table 5. Distribution of Male and Female Employees by Education and Job--Second Finkelstein Hypothetical (Fairness 1 Leads to Unfairness 2)

Education	Male			Female		
	Clerical	Professional	Total	Clerical	Professional	Total
	a. Before Promotion					
High School	15	0	15	30	0	30
College	30	0	30	15	0	15
Total	45	0	45	45	0	45
	b. After Promotion					
High School	10	5	15	20	10	30
College	10	20	30	5	10	15
Total	20	25	45	25	20	45

Table 6. Distribution of Male and Female Employees by Education and Job--Variation on Finklestein's First Hypothetical (Even After Unfairness 1 (Against Women), There Exists Unfairness 2 (Against Men))

	Male			Female		
Education	Clerical	Professional	Total	Clerical	Professional	Total
	a. Before Promotion					
High School	20	0	20	25	0	25
College	25	0	25	20	0	20
Total	45	0	45	45	0	45
	b. After Promotion					
High School	10	10	20	15	10	25
College	5	20	25	5	15	20
Total	15	30	45	20	25	45

Table 7. Distribution of Male and Female Applicants and Hires by With of Without Experience from *Sears* Case

	Total Hires	Female Perc. of Applicants	Female Perc. Hires	Number Female Hires	Number Male Hires
With Experience	50	40.0	20.0	10	40
Without Experience	50	70.0	30.0	15	35
Total	100			25	75

Table 8. Distribution of Male and Female College-Educated Employees as Shown in Table 1, With Hypothetical Female Percentage of Applicants Added

				Number (and Distibution) of Hires	
	Hires	Female Perc. of Applicants	Female Perc. of Hires	Female	Male
Professional	21	20.0	19.0	4 (20.0)	17 (42.5)
Clerical	39	40.0	41.0	16 (80.0)	23 (57.5)
Total	60			20 (100.0)	40 (100.0)

Table 9. Hypothetical Distribution of Male and Female Employees of Similar Skill Level--Employer One

	Hires	Female Perc. of Applicants	Female Perc. of Hires	Number (and Distibution) of Hires: Female	Number (and Distibution) of Hires: Male
High Job	100	Unknown	20.0	20 (33.3)	80 (57.1)
Low Job	100	Unknown	40.0	40 (66.7)	60 (42.9)
Total	200			60 (100.0)	140 (100.0)

Table 10. Hypothetical Distribution of Male and Female Employees of Similar Skill Level--Employer Two

	Hires	Female Perc. of Applicants	Female Perc. of Hires	Number (and Distibution) of Hires: Female	Number (and Distibution) of Hires: Male
High Job	100	Unknown	20.0	20 (20.0)	80 (40.0)
Low Job	200	Unknown	40.0	80 (80.0)	120 (60.0)
Total	300			100 (100.0)	200 (100.0)

Table 11. Hypothetical Distribution of Male and Female Employees of Similar Skill Level--Employer Three

	Hires	Female Perc. of Applicants	Female Perc. of Hires	Number (and Distibution) of Hires: Female	Number (and Distibution) of Hires: Male
High Job	100	Unknown	10.0	10 (33.3)	90 (52.9)
Low Job	100	Unknown	20.0	20 (66.7)	80 (47.1)
Total	200			30 (100.0)	170 (100.0)

Table 12. Hypothetical Distribution of Male and Female Employees of Similar Skill Level--Employer Four

	Hires	Female Perc. of Applicants	Female Perc. of Hires	Number (and Distibution) of Hires: Female	Number (and Distibution) of Hires: Male
High Job	100	Unknown	10.0	10 (50.0)	90 (50.0)
Low Job	100	Unknown	10.0	10 (50.0)	90 (50.0)
Total	200			20 (100.0)	180 (100.0)